CODIB-D-58/2
16 June 1960

UNITED STATES INTELLIGENCE BOARD

COMMITTEE ON DOCUMENTATION

Intelligence Community Position
on the IBM Proposal for an
Automatic Translation Facility

1. I am circulating CODIB-D-58, reorganized as to arrangement of paragraphs, but with only minor changes in paras. 2, 3 and 5 (old paras. 2, 3 and 8). Added, however, is para. 10, the Conclusion, "That it would be premature to accept the IBM proposal or any appreciable portion thereof."

2. With the exception of some components of the Air Force, I believe that the views expressed at our last meeting by CODIB and SCOMT representatives support this conclusion.

3. As to the first nine paragraphs, it seems to me that they can pretty well stand as written. The first four are general propositions and indicate goals. Need for goals was pointed out by Dr. Alexander, though not necessarily as spelled out here. The next five paragraphs were not in my opinion seriously contradicted, except in some major respects by Mr. Samson. I believe that they support the conclusion reached.

4. Please let me know whether you can go along with this draft or in what respects you consider it still deficient. If agreement can be reached, it should be a useful document. If agreement cannot be reached, the CODIB position will be taken to be limited to what is in the minutes of the 13 June meeting (CODIB-M-23) and the Conclusion set forth in para. 10. In the latter case, no further draft will be circulated.

[redacted] 25X1

Paul A. Borel
Chairman

COMIB-D-58/2
Second Draft
16 June 1960

UNITED STATES INTELLIGENCE BOARD

COMMITTEE ON DOCUMENTATION

Intelligence Community Position on the IBM Proposal for an Automatic Translation Facility

1. Need for real-time translation

The intelligence community has an urgent need for real-time translation. Even if a human translation effort were capable of satisfying the entire volume of intelligence community translation requirements, the time lag (in spite of priority scheduling) and cost inherent in the human effort still would remain prohibitive. The trend toward increasing volume and rates of production of world scientific and technical literature alone would dictate the need for real-time translation.

2. Status of the U.S. MT program

The majority of the machine translation efforts in the U.S. are directed toward long-range research goals. Most projects are engaged

25X1

[redacted]

results indicative of early operational capability.

3. Degree of support justified

The degree of achievement in MT to date, and the requirements for translations (including crude translations available at an early date), justify USIB support for the planning and eventual creation of a central machine translation facility as a service of common concern to the U.S. Government and to private organizations as well. The initial goal or any program should be a pilot operation with production coverage limited both as to languages and disciplines, but where possible including experimentation in extracting, abstracting and indexing. USIB support should be augmented year to year, to the extent justified by an evaluation of results, until a capability has been achieved for translating all languages and disciplines required by the USIB.

4. The extent of departmental participation

The extent of financial support generally committed by members of USIB and other departments or agencies of Government in an over-all program, including the development phase of a central machine translation facility, should be a function of their responsibility to promote research, as in the case of the National Science Foundation, or their need for producing human translations, as in the case of the Central Intelligence Agency.

5. Current status of IBM project

The IBM technique is machine and dictionary oriented, that is, a special purpose machine with an extremely large-capacity storage has

been developed to accommodate a method (not yet completely formulated) of translating language by machine. At the present stage of development, entry-for-entry Russian-to-English translation which, in most cases, is readable but not necessarily accurate has been accomplished. Programming, in the usual sense of digital machines, is not required and the entire series of operations effecting translation is reduced to a table look-up procedure. The current basic machine configuration

25X1

[redacted]

tially, translation is accomplished by dictionary look-up on a single pass with certain address modifications to accommodate local context.

6. Problem areas and proposed solutions

The fact that the approach is basically lexicographic in nature leaves a number of essentially important language problems unresolved; e.g., certain local context problems, proper nouns, word order, article insertion, synthesis of prepositional structure, syntax and semantic analysis. The accuracy of translation is contingent upon the solution of these problems. Lack of programming capability which limits or excludes the range of logical operations required in the solution of syntax and semantic problems immediately imposes a constraint on the system's intrinsic power of translation. IBM proposes to modify the basic systems configuration by including a "word analyzer scheme" designed to resolve the basic problems of syntax and meaning. Optimization of input and output will be accomplished by phasing-in a

series of engineering developments (input: flexowriter tape, steno-writer, Baird-Atomic print reader; output: flexowriter, CRT photo printer, high-speed impact printer) now in progress. The rate at which the quality of the product (all languages indicated) will improve to and beyond the point of acceptability as a result of these changes cannot be estimated with any degree of reasonable certainty. A prudent expectation would make it well beyond the three year period of time stipulated in the proposal.

7. Validity of the IBM technique

The IBM technique in theory is a valid approach to the machine translation problem; however, because of its present stage of development, its potential as a practical translation system per se cannot be reasonably estimated or evaluated. The proposed modification of the basic systems configuration can only be successful to the extent that it incorporates the theoretical solutions to the syntactic and semantic problems.

8. Other applications of special-purpose equipment

The large-capacity photoscopic memory and the equipment associated with it hold forth extremely interesting prospects for language data processing other than machine translation. The system is especially adopted to the manipulation of non-numerical data. As such, its use in information storage and retrieval, automatic abstracting and indexing appears to be not only feasible but also distinctly advantageous over general-purpose digital computers.

9. Relationship of IBM proposal and other MT programs

The IBM system, after considerable hardware development, now requires the expenditure of vast funds for a period of intensive research and development on a method of translating language by machine. The development of such a method depends essentially on the success in solving the semantic and syntactic problems not only by IBM research but also by other MT groups. A stored-program general-purpose digital computer allows for a wider range of sophisticated logical operations which are essential to the solution of syntactic-semantic problems. With recent and anticipated advances in storage technique for digital computers, storage will no longer be a problem capacity-and-access-time-wise.

10. Conclusion

The Committee on Documentation, based upon a review of the IBM proposal, a visit to the IBM Research Laboratory (Mohansic), Air Force/ARDC plans over the next 18 months to prove the operational feasibility of its MT program, has concluded:

> That it would be premature to accept the IBM proposal or any appreciable portion thereof.